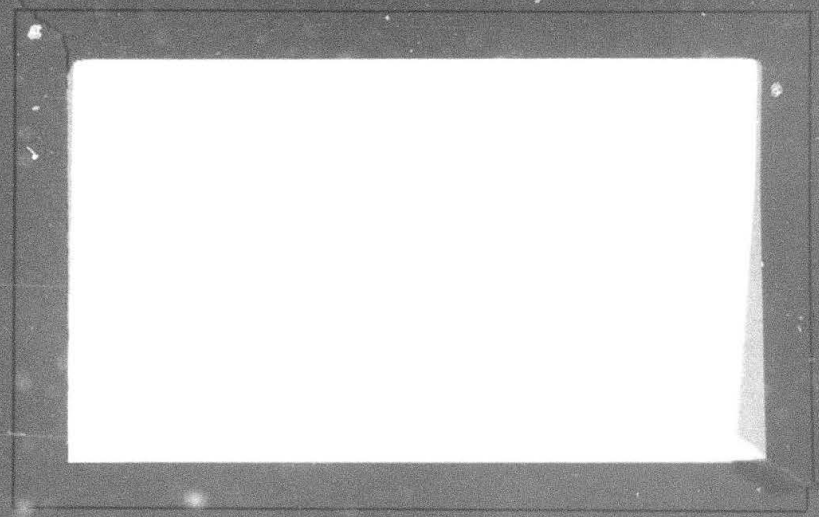
AD A129353



— STATISTICS —

— OPERATIONS RESEARCH —

— MATHEMATICS —

DTIC
ELECTE
JUN 15 1983
S D
A

DESMATICS, INC.

P.O. Box 618
State College, PA 16801

DTIC FILE COPY

This document has been approved for public release and sale; its distribution is unlimited

83 06 14 048

DESMATICS, INC.

P. O. Box 618
State College, PA. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

STATISTICAL IMPACT ACCELERATION INJURY PREDICTION MODELS BASED ON $-G_x$ ACCELERATOR DATA AND INITIAL HEAD CONDITIONS

by

Kevin C. Burns
and
Dennis E. Smith

TECHNICAL REPORT NO. 112-13

May 1983

DTIC
ELECTE
JUN 15 1983
D
A

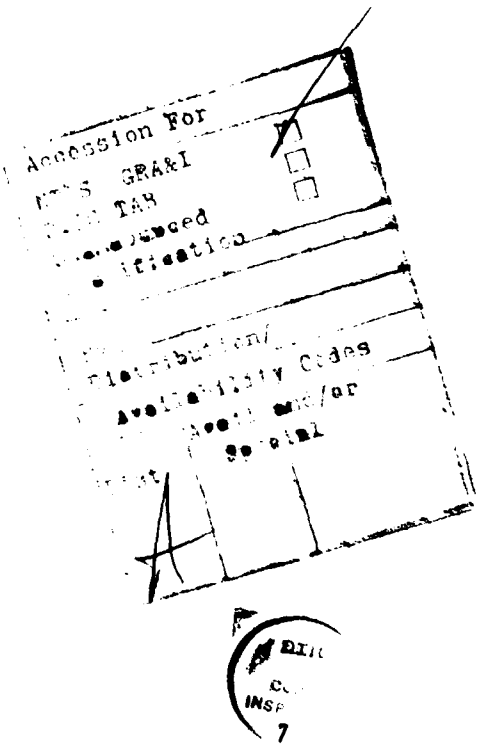
This study was supported by the Office of Naval Research
under Contract No. N00014-79-C-0128, Task No. NR 207-037

Reproduction in whole or in part is permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	MODEL CONSTRUCTION	4
III.	COMPARISON WITH EARLIER MODELS	8
IV.	SUMMARY	15
V.	REFERENCES	17



I. INTRODUCTION

Several previous Desmatics technical reports have investigated the use of a logistic function in the development of indirect impact acceleration injury prediction models. Empirical data have been used to fit these models and evaluate their ability to accurately predict fatal injuries. The most recent of these reports [3] explored the possibility of developing a model based on head dynamic response data. Models using only sled acceleration profile terms were also considered, as were models using a mixture of the two types of variables. The results of that study have been influential in the selection of variables to be considered in this investigation.

The prediction models used in this report are of the same functional type as considered previously:

$$P(\underline{x}) = \{ 1 + \exp[- (\beta_0 + \sum_{i=1}^{k} \beta_i x_i)] \}^{-1}$$

where:

$\underline{x} = (x_1, x_2, \ldots, x_k)$ denotes the set of independent variables,

$(\beta_0, \beta_1, \ldots, \beta_k)$ denotes a set of parameter values,

and $P(\underline{x})$ denotes the true probability of injury corresponding to $\underline{x}$. A discussion of how this type of model may be related to the forces and moments in the region of injury may be found in [1].

The generic type of model described above has previously been applied to a set of 63 $-G_x$ accelerator runs involving Rhesus monkeys with securely restrained torsos and unrestrained heads. The data were collected by the Naval Biodynamics Laboratory (NBDL) as part of its research effort on impact acceleration injury prevention. Three sled profile variables and twelve head dynamic response variables were considered for inclusion in the models.

In the models which included both types of variables, peak sled acceleration was found to be the most important variable while the peak z-component of head angular velocity also contributed to the explanatory power of the model. When only head dynamic response variables were used, the peak y-component of head angular acceleration was also found to be important.

Research to date has found that peak sled acceleration is a better predictor of fatality in $-G_x$ Rhesus runs than is any combination of head dynamic response variables yet considered. Furthermore, inertial data are not available for several high level (>85 G) experiments. Since the most useful data for the development of a predictive model are those obtained near the boundary between injury and noninjury, the high level experiments are of primary interest in the formulation of the logistic models considered here. This combination of factors leads to the consideration of models which do not rely on inertial data.

Although head dynamic response data are not available for many of the accelerator runs of greatest interest, there are photographic data available from which the initial yaw and pitch angles of the head may be determined. As mentioned in [1], the initial yaw angle should be highly correlated with the z-component of head angular velocity. It also seems reasonable that the initial pitch angle might be correlated with the y-component of angular acceleration. Thus, inclusion of these initial head conditions should at least partially compensate for the lack of inertial data.

The data base used in this investigation consists of 23 high level $-G_x$ experiments. For Run No. LX0657 the initial pitch angle is unavailable and for Run No. LX3010 the initial yaw angle is not known. Thus, only 21 data vectors can be used in the analysis. Thirteen of these sled runs have been used in previous analyses while the remaining eight experiments could

not be used previously owing to the lack of inertial data. The independent variables used in this analysis are peak sled acceleration (G), initial yaw angle of the head (deg), and initial pitch angle of the head (deg). Because of symmetry about the mid-sagittal plane in $-G_x$ accelerator runs, the absolute value of the initial yaw angle was the variable actually used in the models. As in previous studies, the dependent variable is binary (1 = fatal run, 0 = non-fatal run). A listing of the data set is given in Table 1.

Run Number	Subject Number	Observed Probability	$\hat{P}(x_1,x_2)$	$\hat{P}(x_1,x_2,x_3)$	Peak Sled Acceleration	Initial Yaw	Initial Pitch
LX0660	A03146	0	.0959	.0755	107	0°	40°
LX3188	AR8863	0	.2022	.3086	104	-30°	-10°
LX1362	A03935	0	.3175	.2157	108	30°	40°
LX1894	A03933	0	.3175	.3241	108	-30°	15°
LX1359	A04099	0	.3866	.3699	110	-30°	20°
LX1893	A03924	0	.3866	.5335	110	30°	-10°
LX3015	AR0764	0	.5543	.5525	97	90°	-10°
LX3032	AR3936	0	.6200	.4682	125	0°	60°
LX3709	AR8790	1	.0776	.1047	88	60°	-15°
LX3192	AR8866	1	.5265	.3819	105	60°	30°
LX1363	A03935	1	.5463	.5009	123	0°	40°
LX1905	A04101	1	.6550	.6673	126	0°	30°
LX3031	AR4115	1	.6885	.7661	127	0°	15°
LX1896	A03946	1	.8023	.8726	131	0°	10°
LX1365	A03921	1	.8850	.9101	109	90°	-20°
LX3026	A03923	1	.9691	.9793	127	-60°	-10°
LX3016	AR0764	1	.9868	.9884	124	90°	-10°
LX1360	A04099	1	.9928	.9913	128	90°	5°
LX1895	A03951	1	.9954	.9965	131	-90°	-15°
LX0661	A03146	1	.9959	.9959	158	0°	40°
LX3033	AR3936	1	.9981	.9971	163	0°	60°

Table 1: The Data Set With Predicted Probabilities of Fatality For Two Models

II. MODEL CONSTRUCTION

As mentioned in the previous section, the data base used in this investigation consists of 21 complete observations. Only 16 animals were used for this set of experiments, so some dependence exists in the data as a result of having multiple sled runs with the same animal. As stated in [3], this type of dependence will tend to make the model conservative; i.e., the probability of injury will be overpredicted for a specified $\underline{x}$.

Initially, a model was considered which included the three independent variables and all of their two-way interactions. A backward selection method was employed in order to eliminate those terms which did not contribute significantly to the model's ability to explain the occurrence or nonoccurrence of fatal injury. In this way, all of the interactions were found to be unimportant terms in the model.

Denote the independent variables by x_1, x_2, and x_3, where:

x_1 = peak sled acceleration,

x_2 = absolute value of the initial yaw angle of the head,

and x_3 = initial pitch angle of the head.

The predicted probability of a fatality based on these three variables is:

$$\hat{P}(x_1, x_2, x_3) = \{1 + \exp[-(-18.4 + .157\ x_1 + .0352\ x_2 - .0222\ x_3)]\}^{-1}.$$

A comparison between observed probabilities and predicted probabilities for this model is given in Table 1.

The variable which has the least influence on $\hat{P}(x_1, x_2, x_3)$ is x_3, the initial pitch angle. A test of whether x_3 is making a significant contribution to the model can be performed by calculating the following quantities:

L_1 = -2 log likelihood for model containing (x_1, x_2)

and L_2 = -2 log likelihood for model containing (x_1, x_2, x_3).

Under the null hypothesis that x_3 does not improve the model, the statistic $L_1 - L_2$ has an approximate Chi-square distribution with 1 degree of freedom. For this data set, the statistics are given below:

$$L_1 = 17.695\ ,\quad L_2 = 17.213\ ,\quad L_1 - L_2 = 0.482\ .$$

Therefore, there is no significant effect due to x_3 ($p = .487$), given that x_1 and x_2 are already in the model.

Since x_3 does not significantly improve the model's ability to explain the data, it is probably better to use a model based only on x_1 and x_2. The predicted probability of a fatality for this reduced model is:

$$\hat{P}(x_1, x_2) = \{ 1 + \exp[-(-18.5 + .152\ x_1 + .0442\ x_2)] \}^{-1}\ .$$

The predicted probabilities for this model are also given in Table 1. Both x_1 and x_2 are important in this model so no further reduction is possible.

The two models given in this section explain the variation in the data about equally well. However, when using these models to predict new observations, the simpler model is probably more trustworthy. The small effect attributable to the pitch angle of the head, given that peak sled acceleration and initial yaw angle have already been taken into account, could easily be spurious. While any variable added to a model will increase the ability of that model to explain the data, the new variable might be useless as a predictor of future observations. When the perceived effect of a factor is very small, that effect is more likely to be the result of random confounding of that factor with some variable which has not been included in the analysis. The confounding can result in an extremely misleading prediction model. Therefore, of the two models given here, the two-variable model is recommended.

Attention will be restricted to it in the remainder of this report.

The predicted probability of a fatality under the specified model is $\hat{P}(x_1,x_2)$, as given above, where x_1 is the peak sled acceleration and x_2 is the absolute value of the initial yaw angle of the head. The predicted region of fatality is the half plane defined by:

$$.152\ x_1 + .0442\ x_2 \geq c\ .$$

The constant c is determined by the particular predicted probability one wishes to use as a cutoff value. For example, $\hat{P}(x_1,x_2) = .5$ corresponds to $c = 18.5$.

III. COMPARISON WITH EARLIER MODELS

In this section a comparison is made between the model developed in this report and two of the models presented in [3]. As an aid in this comparison, it has been found useful to estimate one other model. This last model includes only peak sled acceleration as an independent variable but is based on the entire set of available $-G_x$ accelerator data. As before, a logistic response model is assumed to be appropriate.

Since the models developed in [3] were based on a different set of data, they cannot be compared directly with the model presented in Section II. However, there are thirteen observations common to the two data bases and some assessment of relative performance may be obtained by comparing the predicted probabilities, using the various models, for this subset of the data. In order to facilitate this comparison, the following notation is defined:

y_1 = absolute value of the peak z-component of head linear acceleration,

y_2 = peak head linear resultant acceleration,

y_3 = absolute value of the peak y-component of head angular acceleration,

and y_4 = absolute value of the peak z-component of head angular velocity.

As before, x_1 and x_2 denote peak sled acceleration and the absolute value of the initial yaw angle of the head, respectively.

In keeping with the notation used earlier in this report, let $\hat{P}(x_1,x_2)$ denote the predicted probability of a fatality for the model given in Section II. Further, consider the best three-variable head dynamic response model, given in [3], which involves the three variables y_1, y_2, and y_3. Denote the predicted probability of a fatality using this model by $\hat{P}(y_1,y_2,y_3)$.

Also consider the best two-variable mixed model, given in [3], which involved x_1 and y_4. Let $\hat{P}(x_1,y_4)$ denote the predicted probability for this model. A comparison of these predicted probabilities for the thirteen common data vectors is given in Table 2. (Minor inaccuracies have been found in the earlier technical report. The predicted probabilities given here are therefore slightly different from those given previously.)

None of the three models compared in Table 2 is a better predictor of fatality for each of the data points under consideration. It is therefore necessary to use some form of summary statistic to aid in the comparison. Perhaps the most obvious measure of performance is given by:

$$Q_1 = \sum_{i=1}^{13} (O_i - \hat{P}_i)^2 ,$$

where O_i is the observed probability of a fatality (0 or 1) for observation i and $\hat{P}_i$ is the corresponding predicted probability. However, use of Q_1 has one serious drawback. Observations with predicted probabilities near .5 automatically contribute large terms to the sum, since the observed probability must be either 0 or 1. It seems reasonable that the statistic used should only be large if an observed fatal run has a predicted probability less than .5 or an observed nonfatal run has a predicted probability greater than .5 . One statistic of this type is:

$$Q_2 = \sum_{i=1}^{13} \frac{(O_i - \hat{P}_i)^2}{\hat{P}_i(1-\hat{P}_i)} .$$

The weights used in this sum are the inverse variances of binomial observations where $\hat{P}_i$ is the probability that the observation is 1.

Q_1 and Q_2 have been calculated for the three models being compared. The corresponding values are:

Run Number	Subject Number	Observed Probability	$\hat{P}(y_1,y_2,y_3)$	$\hat{P}(x_1,y_2)$	$\hat{P}(x_1,x_2)$
LX3188	AR8863	0	.1179	.3241	.2022
LX1362	AO3935	0	.1190	.3889	.3175
LX1894	AO3933	0	.6328	.4996	.3175
LX1359	AO4099	0	.1620	.3326	.3866
LX1893	AO3924	0	.3289	.5324	.3866
LX3709	AR3790	1	.2116	.0800	.0776
LX3192	AR8866	1	.2562	.3180	.5265
LX1363	AO3935	1	.0845	.7345	.5463
LX1905	AO4101	1	.9450	.8081	.6550
LX1896	AO3946	1	.6962	.9028	.8023
LX1365	AO3921	1	1.0000	.8699	.8850
LX1895	AO3951	1	.9999	.9637	.9954
LX1360	AO4099	1	.9106	.9628	.9928

Table 2: Comparison of Predicted Probabilities for Observations Common to the Two Data Sets.

Model	Q_1	Q_2
y_1, y_2, y_3	2.679	20.732
x_1, y_4	2.348	18.328
x_1, x_2	1.994	16.976

Both Q_1 and Q_2 measure, in some sense, lack of fit of the model being considered. That is, they are large when the model is not adequately explaining the observed pattern of fatalities. Thus, using either criterion, $\hat{P}(x_1, x_2)$ performs better for this subset of the data than does either of the models taken from [3]. Also, $\hat{P}(x_1, y_4)$ explains the data better than does $\hat{P}(y_1, y_2, y_3)$, as could be expected from previous results.

There are at least two possible reasons for the fact that $\hat{P}(x_1, x_2)$ performs better than $\hat{P}(x_1, y_4)$. First, it is possible that the initial yaw angle of the head is a better predictor of injury than is the peak z-component of head angular velocity. However, it is also possible that the peak z-component of head angular velocity is the better predictor but that its coefficient in the model is not as well estimated. This circumstance could easily arise as a consequence of the fact that different data sets were used to estimate the two models.

The data set used to estimate the coefficients in $\hat{P}(x_1, y_4)$ consisted of 63 observations while only 21 observations were used to estimate $\hat{P}(x_1, x_2)$. However, those 21 observations were all high-level experiments and thus nearer to the boundary between injury and non-injury. That type of observation is more useful for estimating the coefficients in a logistic response model. It is therefore necessary to quantify the usefulness of each observation in order to determine the relative value of the two data sets. This is usually accomplished by calculating some function of Fisher's information matrix [2].

Suppose w is the dependent variable to be measured. Then for any particular set of values of the independent variables $\underline{x}$, w has a probability distribution which may depend on the vector of parameters $\underline{\beta}$. Denote the probability density function of w at $\underline{x}$ by $p(w|\underline{x},\underline{\beta})$. For a single observation on w at $\underline{x}$, the Fisher information matrix for $\underline{\beta}$ is that matrix whose (i+1,j+1) element is given by:

$$E\{-\partial^2 \log p(w|\underline{x},\underline{\beta})/\partial\beta_i\partial\beta_j\},$$

where $\underline{\beta} = (\beta_0,\beta_1,\ldots,\beta_k)$ and E denotes the expectation over the probability distribution of w. The information matrix of N independent observations taken respectively at $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_N$ is $L(X;\underline{\beta}) = \sum_{i=1}^{N} J(\underline{x}_i;\underline{\beta})$, where $J(\underline{x}_i;\underline{\beta})$ is the information matrix for a single observation, as given above, and $X = (\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_N)$. A more complete discussion of this concept may be found in [2].

For the injury prediction models being considered here, w is binary (1 = fatality, 0 = nonfatality). The probability density function of w, given $\underline{x}$, is:

$$p(w|\underline{x},\underline{\beta}) = [P(\underline{x})]^w[1-P(\underline{x})]^{1-w};\ w = 0,1,$$

where $P(\underline{x}) = \{1+\exp[-\beta_0 + \sum_{i=1}^{k}\beta_i x_i)]\}^{-1}$.

For this model, with $x_0 = 1$, the (i+1,j+1) element of $J(\underline{x};\underline{\beta})$ is:

$$x_i x_j P(\underline{x})[1-P(\underline{x})].$$

Since $J(\underline{x};\underline{\beta})$ depends on $P(\underline{x})$, the information matrix cannot be computed unless the correct model is known. It is possible, however, to obtain some indication of how much information is available in the samples by using the estimate of $P(\underline{x})$. This estimate must, of course, be applicable to both data sets so that only the estimated information matrices will be directly comparable.

The only variable which is available for all of the observations in both

data sets is peak sled acceleration. Therefore, the estimate of $P(\underline{x})$ to be used in computing the information matrices must be a function of this variable alone. Fortunately, peak sled acceleration is known to be a good predictor of fatality. Furthermore, 93 total observations are available for the estimation procedure, so that the coefficients in the model should be fairly well estimated.

The estimated model using only peak sled acceleration as an independent variable is:

$$P(\underline{x}) = \{1 + \exp(13.6 - .120x)\}^{-1}.$$

The predicted values from this model can now be used to estimate $J(\underline{x};\underline{\beta})$ for each of the 93 observations. Three different information matrices have been calculated and are given below. I_1 refers to the complete set of 93 observations, I_2 refers to the 62 observations used to fit models involving head dynamic response variables, and I_3 refers to the set of 21 observations used to estimate the models given earlier in this report.

$$I_1 = \begin{pmatrix} 3.72 & 411.25 \\ 411.25 & 46458.52 \end{pmatrix}$$

$$I_2 = \begin{pmatrix} 2.36 & 259.22 \\ 259.22 & 20019.05 \end{pmatrix}$$

$$I_3 = \begin{pmatrix} 3.13 & 356.87 \\ 356.87 & 41174.91 \end{pmatrix}$$

The inverse of the information matrix provides a lower bound for the variance matrix of an unbiased estimate of $\underline{\beta}$. It is therefore desirable for the inverse to be "small" or, equivalently, for the information matrix to be "large" in some sense. One popular optimality criterion is D-optimality, which says to maximize the determinant of the information matrix. Since $|I_3| = 1521$ and $|I_2| = 1290$, the sample used in this report contains

more information, according to this criterion, than does the sample used in the previous technical report. It is therefore not surprising that the model developed here performs better on the subset common to both samples. However, D-optimality is by no means the only criterion used to evaluate designs. A number of different functions of the information matrix can be used. The only time one matrix can be said to be larger than another is when their difference is positive definite. In this case $I_3 - I_2$ is not positive definite (the determinant is negative) so it is impossible to say definitively which sample contains more information.

IV. SUMMARY

Using a data base of 21 high-level accelerator runs, models have been considered which included peak sled acceleration, initial yaw angle of the head, and initial pitch angle of the head as independent variables. While initial yaw was found to be a useful predictor of injury, the initial pitch angle was found to make little, if any, additional contribution to the model, given that the other variables had already been included. Furthermore, no significant interactions between the three variables were found to exist.

In a previous technical report, [3], models were developed which depended on the use of head dynamic response data. Of the 63 data vectors used in that investigation, thirteen observations included head initial conditions as well. The model developed in this report was compared to the earlier models on the basis of each model's ability to predict injury for the thirteen common observations. Using two different relative performance criteria, the initial conditions model was found to perform better than the others.

Since the model developed in this report used a different set of data than the earlier models, there was some concern that its superior performance on the common data set might be the result of the model being better estimated, rather than reflect the relative worth of the variables. Therefore, an attempt was made to quantify the amount of information available in each sample. Fisher's information matrix was calculated for each sample, using an estimate of the predicted probability for each data point. These estimates were taken from a single-variable model using peak sled acceleration and based on all available $-G_x$ acceleration data. Unfortunately, no clear decision could be made as to which sample contains the most information. Using at least one

criterion, however, the data base used in this report is better. Thus, the model that performs best on the common observations is the model based on:

x_1 - peak sled acceleration

and x_2 - absolute value of initial yaw angle of the head.

However, because at least part of its performance may be due to more inherent information content in the observations on which it was based, there is no conclusive evidence that x_1 and x_2 are the best predictors.

V. REFERENCES

[1] Lustick, Leonard S., "Injury Model for Rhesus", Memorandum From Head, Mathematical Sciences Department To Head, Biomedical Research Department, NBDL, 17 February 1983.

[2] Silvey, S. D., Optimal Design, Monograph on Applied Probability and Statistics, General ed. D. R. Cox, F. R. S. Chapman and Hall, New York, N.Y., 1980.

[3] Smith, Dennis E. and Aarons, David, "Research on the Development of a Statistical Impact Acceleration Injury Prediction Model From $-G_x$ Accelerator Runs," Technical Report No. 112-11, Desmatics, Inc., 1982.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 112-13	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STATISTICAL IMPACT ACCELERATION INJURY PREDICTION MODELS BASED ON $-G_x$ ACCELERATOR DATA AND INITIAL HEAD CONDITIONS		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kevin C. Burns and Dennis E. Smith		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0128
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P.O. Box 618 State College, PA 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 207-037
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE May 1983 13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Impact Acceleration Injury
$-G_x$ Acceleration
Statistical Prediction Model
Fisher's Information Matrix

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Statistical impact acceleration injury prediction models are developed using data from 23 high-level $-G_x$ acceleration runs. These runs involve Rhesus monkeys with securely restrained torsos and unrestrained heads. The models are based on peak sled acceleration and initial head conditions. The model predictions are compared with those given in an earlier report based on different data and an estimate of Fisher's information matrix is used to evaluate the relative worth of the two data bases.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)